

# Modeling User Arguments, Interactions, and Attributes for Stance Prediction in Online Debate Forums

Minghui Qiu \*    Yanchuan Sim †    Noah A. Smith †    Jing Jiang \*

## Abstract

Online debate forums are important social media for people to voice their opinions and debate with each other. Mining user stances or viewpoints from these forums has been a popular research topic. However, most current work does not address an important problem: for a specific issue, there may not be many users participating and expressing their opinions. Despite the sparsity of user stances, users may provide rich side information; for example, users may write arguments to back up their stances, interact with each other, and provide biographical information. In this work, we propose an integrated model to leverage side information. Our proposed method is a regression-based latent factor model which jointly models user arguments, interactions, and attributes. Our method can perform stance prediction for both warm-start and cold-start users. We demonstrate in experiments that our method has promising results on both micro-level and macro-level stance prediction.

## 1 Introduction

Online debate forums are important social media for people to voice their opinions and engage in debates with each other. Mining user stances and viewpoints from these forums has been a popular research area [1; 2; 3; 4; 5; 6]. One potential application is understanding public opinion; e.g., what are the popular stances on the Affordable Care Act, how do they associate with different subpopulations, and how are they changing over time? However, there may be a low online participation rate of Internet users in online discussion forums relating to any particular debate. For example, in our dataset collected from the debating website CreateDebate,<sup>1</sup> where users can explicitly state their stances on user-created debate topics, 278 people participated in the debate titled “Should guns be banned in America,” while only 3 people participated in the debate “ObamaCare.” As a result, if we consider all the registered users and existing debates on CreateDebate, only 0.4% of the full set of user stances are observed in the data. In

this paper, we are interested in *predicting* a user’s stance on a debate topic in which she has not participated. Note that this target user may have not expressed her stances before, i.e., the user may be a cold-start user.

This problem is similar to item recommendation, where a user’s purchase history is used to predict her preference for a new item. Collaborative filtering [7] is a technique commonly used to alleviate the data sparsity problem in item recommendation. Probabilistic matrix factorization (PMF) [8] has been shown to be an effective collaborative filtering algorithm, and we extend PMF in our work.

One notable problem with PMF is that it cannot perform stance prediction for users without any past stances. To alleviate this problem, we incorporate user attributes and user-generated content in online discussions. To be more specific, we incorporate three types of information that are prevalent in online discussions: user arguments that are used to back up their stances, user interactions from texts exchanged between users, and user attributes from their biographical information. Such rich information can help with stance prediction on both warm-start users and cold-start users and provide valuable user and issue profiling. Hence, we propose an integrated model that uses such information to provide insights into user behaviors and issues.

Using data from CreateDebate, we propose a unified model for the task of user stance prediction. Firstly, to incorporate user attributes from their biographies, we use a regression-based latent factorization method [9] to profile users. In this method, each user’s latent factors are aggregated from factors associated with the user’s attributes and user-specific deviations. This setting allows us to profile users who have no past stances, i.e., we can predict stances for cold-start users. We further introduce a novel binomial matrix factorization (BMF) model in the context of categorical ratings (i.e., user stances). This method extends the original PMF [8] model, which is designed for numerical ratings. Furthermore, users write arguments to support their stances, which provide textual cues to understand different topics involved in the issue. Like [10], we associate each latent factor dimension with a topic so as to produce an

<sup>\*</sup>School of Information Systems, Singapore Management University

<sup>†</sup>Language Technologies Institute, Carnegie Mellon University

<sup>1</sup><http://www.createdebate.com>

intuitive explanation for the hidden factors from BMF and improve stance prediction results. In addition, we find that incorporating features about a user’s interaction network provides us with a way to infer relationships between users, which we can leverage to better predict user stances. To infer the model parameters, we employ Monte Carlo EM [11; 12] and adapt a fast inference method based on SparseLDA [13].

This work also aims to contribute to the problem of inferring public opinion from freely available social media text and metadata [14; 15; 16]. Such approaches have the potential to complement traditional surveys and polls. We focus on debate forums with rich user-contributed texts, opinions, and interactions on diverse topics. We formulate the task around predicting held-out user stances. Although online forums have been explored in the past for questions such as mining user stances [2; 3; 5; 6], detecting subgroups in online communities [17; 18; 19; 20], identifying user interactions [21; 22; 23; 24; 25], knowledge discovery [26], and detecting new and significant developments in science and technology [27], to the best of our knowledge, this is the first study on both micro-level and macro-level stance prediction in a debate forum leveraging rich user metadata. We find our methods tend to agree with polls from Gallup, despite the fact that aggregated user stances are different from Gallup. This is promising as the polls are based on survey data collected through a labor-intensive procedure, while our method could serve as a cheap and effective way to complement them.

Our contributions are as follows:

- We propose a regression-based latent factor model which jointly models user arguments, interactions, and attributes for stance prediction.
- We study a fast inference method for the model.
- Our experiments show promising results on micro-level stance prediction for both warm-start and cold-start users.
- Our experiments show that our model has a good result on macro-level stance prediction, which shows the potential to complement traditional surveys and polls.

## 2 Problem Definition

In Table 1, we present an excerpt of user arguments from a debate page in CreateDebate. In CreateDebate, each debate issue  $i$  focuses on a particular debate question, for example, “Does God exist?” Each debate issue has defined stances, which are usually “Yes” and “No” stances for the issue. In addition, each issue  $i$  has a set of threaded arguments, where each argument can be

an independent post or a reply to an earlier argument. Each argument is authored by a user  $u$ , and explicitly contains his stance  $r_{u,i}$  on the particular issue, e.g., user A in our example takes the “Yes” stance. One user can write multiple arguments on an issue. We represent the text of the  $n$ th argument from user  $u$  on issue  $i$  using a bag of bigrams  $w_{u,i,n}$ .<sup>2</sup> If the  $n$ th argument by user  $u$  is a reply/interaction post, the user  $u$  needs to specify whether she wants to “dispute,” “support,” or “clarify” the recipient post. We take advantage of this metadata using an interaction polarity  $l_{u,i,n} \in \{\text{positive, negative}\}$ , and assign  $l_{u,i,n}$  to be “negative” when the user’s argument disputes an earlier post and “positive” otherwise.

Debate: <i>Does God Exist?</i>	
“Yes” stance	“No” stance
<p><b>User A:</b> Theists: I believe God exists. Atheists: I believe God doesn’t exist. Both rely on belief ... (“Yes” stance) ⊙<b>User B (Dispute):</b> Whoops. wrong. more like “I don’t believe in God.” Both rely on belief. Thing about theists is they look beyond what they can see... (“No” stance) ⊙<b>User A (Dispute):</b> ...Athiesm relies on belief. You BELIEVE there’s no God ... (“Yes” stance)</p>	<p><b>User J:</b> If there is no evidence leading up to a God, I dont believe... (“No” stance) ⊙<b>User K (Disputed):</b> ...how can we prove that it doesn’t exist??? We have no choice but to accept it (“Yes” stance) ⊙<b>User L (Disputed):</b> So how can you argue for something that you cannot even interact with on a comparable level? (“No” stance) ⊙<b>User M (Supported):</b> ...Answer: God told the people to kill muslims ... (“No” stance)</p>

Table 1: Sample arguments on the debate “Does God Exist?”

In April 2013, we crawled all arguments of two-sided debates from all 14 categories in the CreateDebate website. For each participant, we also collect these six types of attributes from his biographical information: *party* (e.g., republican, democrat), *religion* (e.g., Catholic, Christian), *gender* (e.g., male, female), *status* (e.g., single, married), *education* (e.g., in college, post-grad.), and *country* (e.g., U.S., Singapore). We leave other attributes like “age” and user self-description in biography as further study. Table 2 displays some statistics on the CreateDebate corpus. We find user-stance and interaction information are sparse in our data.

Our task is to predict a given user  $u$ ’s stance on a target issue  $i$  when the user has not expressed his stance on that issue. We refer to this as micro-level stance prediction. In § 5.2, we also consider macro-level stance prediction, where we estimate the percentage of users holding a certain stance for a particular issue  $i$ .

<sup>2</sup>In this work, we use only a bigram representation. Eschewing unigrams has been shown to provide for more human interpretable terms without hurting task performance [28].

# users	4,994
# issues	1,727
# arguments	39,549 (average 23 per issue)
# unigram tokens	55,308
# unique bigrams	154,724 (after pruning)
# user stances	17,843 (0.21% density)
# interaction links	23,935 (0.12% density based on a symmetric matrix, excluding self-links)
# user attributes	party (1430), religion (926), gender (1430), status (1432), education (835), country (1432)

Table 2: Statistics of the dataset. Bigrams containing stop words or punctuations are removed during pruning.

### 3 Model

We approach the problem using a probabilistic graphical model. The graphical representation of the model can be found in Figure 1.

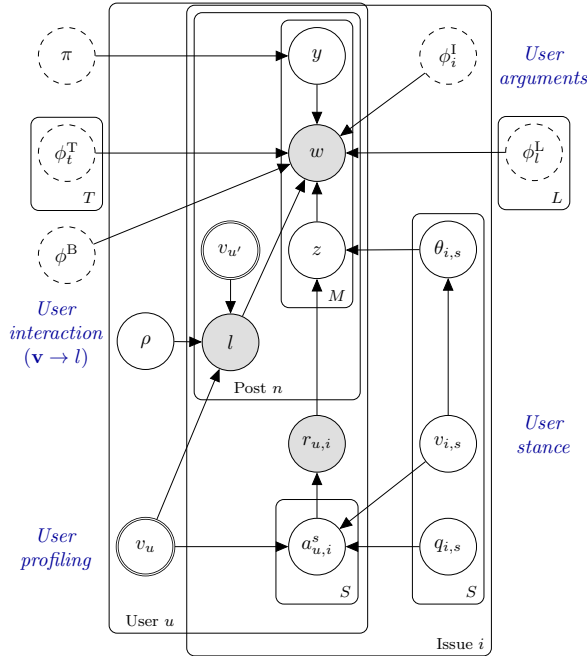


Figure 1: Plate notation for our model. The dashed variables will be collapsed out during Gibbs sampling.  $\rho = \{c_1, c_2, q_u, q_{u'}\}$ , representing two parameters used in user interaction modeling (equation 3.4) and two biases specific to a user and her recipient  $u'$ .  $v_u, v_{u'}, q_u$  and  $q_{u'}$  are fixed by a regression-based latent factorization method, detailed in § 3.1. Hyperparameters are omitted for clarity.

The model is composed of four parts: *i.*) *user profiling*, which considers a regression-based latent factorization method to incorporate user attributes for profiling users; *ii.*) *user stance*, which contains a binomial matrix factorization method for modeling categorical

stance data; *iii.*) *user arguments*, which incorporate textual cues in threaded posts; and *iv.*) *user interaction*, which integrates the positive and negative interaction attributes between users.

**3.1 User Profiling.** Inspired by the work in [9; 29; 30], we consider a regression-based latent factorization method for profiling users. Let  $f_u \in \mathbb{R}^{P \times 1}$  denote user  $u$ 's attributes. We use a binary representation, where each dimension of  $f_u$  is set as 1 if the corresponding attribute is present in user  $u$ , and 0 otherwise. In this study, we consider user attributes in these six types: *party, religion, gender, status, education, and country.*

We model user latent factors  $v_u$  and bias  $q_u$  as:

$$(3.1) \quad \begin{aligned} v_u &= \frac{1}{Z_u} G^\top f_u + \delta_u, \\ q_u &= \frac{1}{Z_u} g^\top f_u + b_u, \end{aligned}$$

where  $G \in \mathbb{R}^{P \times F}$  is a regression coefficient matrix,  $g \in \mathbb{R}^{P \times 1}$  is a regression coefficient vector,  $\delta_u \in \mathbb{R}^{F \times 1}$  is user-specific deviation,  $Z_u$  refers to the number of feature types user  $u$  has provided, i.e.  $Z_u = \sum_p f_{u,p}$ .  $b_u$  is user-specific bias. We pose zero-mean Gaussian priors on  $G, g, \delta_u$ , and  $b_u$ .

**3.2 User Stance.** To model issues, we embed a factor vector  $v_{i,s} \in \mathbb{R}^{F \times 1}$  associated with each stance  $s$  of each issue  $i$ . The factor vector is drawn from zero-mean spherical Gaussian priors,

$$v_{i,s} \sim \mathcal{N}(0, \sigma_i^2 I),$$

where hyperparameters  $\sigma_i^2$  are issue-related variances. This differs from probabilistic matrix factorization in associating multiple factor vectors with a single issue. In this paper, each issue corresponds to two vectors denoting the *support* and *oppose* stances; more stances could also be incorporated here.

Every user  $u$  has an affinity score on each stance  $s$  of an issue  $i$ ,

$$(3.2) \quad a_{u,i}^s = v_u^\top v_{i,s} + q_{i,s},$$

where  $q_{i,s}$  is item-stance bias which is drawn from zero-mean Gaussian priors  $\mathcal{N}(0, \sigma_q^2)$ .

Using a logit function, the probability of user  $u$  choosing a stance  $s$  on issue  $i$  is

$$(3.3) \quad p(r_{u,i} = s) \propto \exp\{a_{u,i}^s\}.$$

This approach models the categorical rating data (stance) in debates. It captures the intuition that a user chooses a stance based on her own preference

over different stances. For example, in the debate “Do you support Obama or Romney in the presidential election?”, a user tends to choose a stance “Obama” when her affinity score of “Obama” is higher than that of “Romney,” i.e.,  $a_{u,i}^{\text{Obama}} > a_{u,i}^{\text{Romney}}$ .

We refer to this way of modeling user stance as binomial matrix factorization (BMF) as it extends probabilistic matrix factorization to two-sided stance data. This framework can be easily extended to more than two sides.

**3.3 User Arguments.** We use a latent Dirichlet allocation topic model [31] to reduce the dimensionality of the text, and combine text data with latent factors from the user stance matrix, grounding each dimension of the hidden factor using inferred topics. Particularly, in our model, topics are in the same space as hidden factors, which is similar to the setting in [10].

We assume a stance-specific topic mixture  $\theta_{i,s}$  for each stance  $s$  of an issue. The reason is that, for each issue, users with different stances tend to have different topic preferences [5].  $\theta_{i,s}$  denotes the relative log-odds of the different topics in issue  $i$  and stance  $s$ , encoding the distribution of topics that are likely to occur when arguing for that particular issue-stance. Specifically,  $\theta_{i,s}^T \propto \exp(v_{i,s}^T)$ , where  $v_{i,s}$  denotes the hidden factors associated with issue  $i$  and stance  $s$ .

The advantages of associating topic distributions with latent factors are two-fold. Firstly, the learnt topics provide an interpretation for factors, as each latent factor dimension is associated with a topic-specific word distribution. Secondly, this helps to reduce ambiguity for latent factors. In the BMF,  $v_u$  and  $v_{i,s}$  can be replaced by  $Uv_u$  and  $Uv_{i,s}$  when  $U^T U = 1$ , as  $v_u^T v_{i,s} = (Uv_u)^T (Uv_{i,s})$ . This means the factors may change considerably while leaving the underlying model unchanged. With this association, the  $\theta$  learnt from texts will regularize the latent factors.

Moreover, we provide a fine-grained categorization of terms (bigrams), where we assume the terms in a user’s argument are drawn from one of the following four term distributions.

- Background term distribution  $\phi^B$ . These are words uniformly distributed in many issues. For example, “united states,” “no longer,” and “things like.”
- Issue-specific term distributions  $\phi_i^I$ . Words that are related to the debate issue, e.g.: “God existence,” “believe God” for the issue “Do you believe in God?”
- Topical term distributions  $\phi_t^T$  for each topic  $t$  ( $1 \leq t \leq T$ ). For example, terms like “health care,” “federal government,” and “tax cuts” are closely related to the topic “health care” and thus tend to

have high probabilities under this topic.

- Interaction term distribution  $\phi_l^I$  for each type of interaction  $l$ . These are words related to “positive” and “negative” interactions, for example, “i agree,” “good point” for “positive” interaction, and “not agree,” “not like” for “negative” interactions. In our work, the interaction polarity of an argument is observed, and this information is fed into our model to learn those interaction terms. Recent work in [22] also considers modeling interaction terms, however, they assume the interaction polarity is not available and use a maxent component to guide the model to find interaction terms.

Additionally, we incorporate switching variables  $y$  to decide from which term distribution a bigram is drawn [32; 33].

The generative story of our model of user arguments is

- Draw switching variable type distribution  $\pi \sim \text{Dirichlet}(\gamma)$ .
- Draw  $\phi^B \sim \text{Dirichlet}(\eta^B)$ .
- $\forall$  interactions  $l$ , draw  $\phi_l^I \sim \text{Dirichlet}(\eta^I)$ .
- $\forall$  topics  $t$ , draw  $\phi_t^T \sim \text{Dirichlet}(\eta^T)$ .
- $\forall$  issues  $i$ , draw  $\phi_i^I \sim \text{Dirichlet}(\eta^I)$ .
- $\forall$  stances  $s$  in issue  $i$ , the issue-stance topic distribution is determined by a logit function:  $\theta_{i,s}^t \propto \exp(v_{i,s}^t)$ .
- $\forall$  terms  $w$  in the  $m$ th position of argument  $n$  from user  $u$  on issue  $i$ :
  - Draw switch  $y_{u,i,n,m} \sim \text{Discrete}(\pi)$ .
  - Draw topic  $z_{u,i,n,m} \sim \text{Discrete}(\theta_{i,s})$ .
  - Draw term  $w$ ,

$$w \sim \begin{cases} \text{Discrete}(\phi^B) & \text{if } y = B, \\ \text{Discrete}(\phi_i^I) & \text{if } y = I, \\ \text{Discrete}(\phi_{z_{u,i,n,m}}^T) & \text{if } y = T, \\ \text{Discrete}(\phi_{u,i,n}^I) & \text{if } y = L. \end{cases}$$

**3.4 User Interaction.** In CreateDebate data, user interactions are observed. As illustrated in Table 3, more positive user interactions are observed within difference stances. We thus use it to inform our model in stance prediction, i.e., to guide the model to predict a user’s stance to be the same with other users with whom she has positive interactions, and different from those with whom she has negative interactions.

This motivates us to associate the similarity of users in the latent factors with the polarity of user interactions. To measure the similarity of users in the latent factors, we simply use the inner product of user factors with user biases. We leave other alternatives as future work. We then enforce a high probability

	Same stance	Different stance
Positive interactions	2,677 (71%)	1,101 (29%)
Negative interactions	932 (26%)	2,691 (74%)

Table 3: Confusion matrix for positive/negative interaction user pairs vs. user pairs with same/different stances. Interactions between users are aggregated across all in issues in our corpus.

of observing a positive interaction polarity for users who are highly similar. Specifically, let  $u'$  denote the recipient of user  $u$ 's  $n$ th post in issue  $i$ , we sample user interaction polarity  $l_{u,i,n}$  as:

$$p(l_{u,i,n} = +) = \mathbb{S}\left(c_1(v_u^\top v_{u'} + q_u + q_{u'}) - c_2\right),$$

$$(3.4) \quad p(l_{u,i,n} = -) = 1 - p(l_{u,i,n} = +),$$

where  $\mathbb{S}(\cdot)$  is logistic function,  $c_1 \sim \mathcal{N}(1, \sigma^2)$  which encourages a positive value, and  $c_2$  is also sampled from a Gaussian distribution with zero mean, i.e.,  $c_2 \sim \mathcal{N}(0, \sigma^2)$ ;  $q_u$  and  $q_{u'}$  are user-specific biases, sampled from zero-mean Gaussian distribution  $\mathcal{N}(0, \sigma_q^2)$ ;  $v_u, v_{u'}, q_u$ , and  $q_{u'}$  are fixed by equation 3.1.

#### 4 Inference and Learning

Our goal is to learn the hidden factor vectors and topics of the textual content to accurately model user stances and maximize the probability of generating the textual content. Hence our objective function is to minimize:

$$J = - \sum_{u,i,n} \left( \log p(r_{u,i} | \rho_{u,i,n}) + \log p(l_{u,i,n} | \rho_{u,i,n}) \right. \\ \left. + \log p(\rho_{u,i,n} | \Upsilon) + \log p(w_{u,i,n} | l_{u,i,n}, v_{i,s}, \Omega) \right),$$

where  $u, i, n$  are user, issue and argument index respectively.  $\rho_{u,i,n} = \{v_i, q_i, G, g, \delta_u, \delta_{u'}, b_u, b_{u'}, c_1, c_2\}$  refers to the set of latent variables related to user  $u$ , recipient  $u'$  of the  $n$ th post of user  $u$ , and issue  $i$ , and  $\Upsilon$  is the set of Gaussian priors for all the variables in  $\rho_{u,i,n}$ .  $\Omega$  denotes all the Dirichlet prior hyperparameters for  $\phi$ . The first three terms denote the probability of generating user stance and interaction given the priors  $\Upsilon$ , where the variable in  $\rho_{u,i,n}$  are to be optimized to minimize the objective function. The last term denotes the probability of observing the text conditioned on  $\theta_{i,s}$  from learnt vector  $v_{i,s}$ , interaction  $l_{u,i}$ , and Dirichlet priors  $\Omega$ .

Exact inference under the posterior distribution is intractable. We use Monte Carlo EM [11; 12], an inference method that alternates between collapsed Gibbs sampling [34] and gradient descent, to estimate parameters in the model. In the E-Step, we perform Gibbs sampling for variables  $\{y, z\}$ , fixing the values

of  $\rho$ . In the M-step, we perform gradient descent to update latent variables in  $\rho$ , fixing the values of  $\{y, z\}$ . Note that the E-step usually takes more time than the M-step. To speed up the E-step, we borrow the idea from inference process in SparseLDA and find that it performs three times as fast as the original setting. For space, we leave all the detailed derivations to the supplementary pages<sup>3</sup>. Note that to further scale up our model for huge data sets, one may resort to use parallel frameworks like [35] and new sampling techniques like [36]. We leave this to future work.

We ran 1,000 iterations of Monte Carlo EM. For Gibbs sampling steps, we ran 400 iterations for burn-in and sampled every 10 iterations to reduce auto-correlation. We fixed the number of topics and dimension of the latent factors at  $T = F = 20$ . (We considered this number from 10, 20, ..., 90, 100 and found the resulting topics to be more meaning at around 20 by manual examination.) For our models and competing baselines, we use grid search on a development set to select the model hyperparameters.

## 5 Experiments

Recall that our task is to predict a user's stance on an issue that she has not commented on. This problem setting is different from existing studies on stance prediction (e.g., [4; 2; 3]) where a user's arguments about an issue are observed but not her stance, and hence incomparable. We design experiments to: (i.) quantitatively evaluate our model with baselines on the tasks of micro-level stance prediction on warm-start and cold-start users; (ii.) examine our model on macro-level stance prediction; (iii.) analyze the efficiency of our inference method; and (iv.) qualitatively examine the learnt term distribution of topics and issues. We leave the study of efficiency of the inference method and the qualitative analysis to supplementary pages.

**5.1 Micro-Level Stance Prediction.** We evaluate on the task of micro-level stance prediction, i.e., predicting user stances on a given issue using learnt user and issue factor vectors and the user interaction network. We perform 10-fold cross-validation on our dataset. Each time, we hold out as a test set 10% of the observed user-issue pairs, i.e., observed user stance on an issue. For each test set, if the issue does not appear in the training set (i.e., a cold-start issue), we will put it back into the training set. As for the rest test sets, we split it into two datasets: one is for warm-start users who have their past stances in the training set, and the other for cold-start user who have no past stances in the training

<sup>3</sup>[www.mysmu.edu/faculty/jingjiang/papers/SDM'15-sup.pdf](http://www.mysmu.edu/faculty/jingjiang/papers/SDM'15-sup.pdf)

set. We remove all the text associated with user-issue pairs in the test set, and the prediction is based solely on users and issues factor vectors learnt from the training set. This setting mimics a real world scenario where a user does not have any prior stance on an issue, but the user has expressed stances on other issues and the issue has other users expressing their stances on it.

**5.1.1 User Attributes.** We first examine the importance of different user attributes for stance prediction task. We use prediction accuracy (Acc.) to measure model performance:  $\text{Acc.} = \frac{1}{|\mathcal{R}|} \sum_{u,i} \mathbb{I}(\hat{r}_{u,i} = r_{u,i})$ .

We refer our base model without any user attributes as BMF-AI, binomial matrix factorization with user arguments and interactions. We evaluate the results by first considering only one type of attributes.

	BMF-AI	+ P	+ R	+ G	+ S	+ E	+ C
Acc.	0.703	<b>0.707</b>	0.705	0.651	0.654	0.698	0.662
s.d.	0.008	0.008	0.007	0.007	0.008	0.007	0.008

Table 4: Micro-level stance prediction results on warm-start users by only incorporating one type of user attributes, averaged across ten folds. P, R, G, S, E, C stand for party, religion, gender, status, education, and country respectively. s.d. refers to standard deviation.

Table 4 shows that, of those considered, only those attributes related to ‘‘ideology’’ (party and religion) are helpful for the task of stance prediction. Furthermore, if we incorporate both party and religion attributes into the model, Acc. rises to 0.712. In the following experiments, we will only incorporate these two attributes.

**5.1.2 Warm-start Users.** We evaluate the following competing models for comparison:

- MB: majority baseline. For each test issue, we predict a user’s stance based on the majority stance on the issue from the training data. This method performs well when the stances are unbalanced, i.e., when an issue has a dominant stance.
- PMF: probabilistic matrix factorization [8]. The original model is designed for numerical ratings. We randomly map one stance of an issue to 0 and the other to 1.
- BMF: binomial matrix factorization. This model differs from PMF in that it assumes a rating for each stance of an issue and draws a stance based on a logit function over stance-specific ratings.
- HFT: hidden factors as topics [10]. Based on PMF, this model further considers user arguments.

- Our model and its variants: BMF-A: BMF with user arguments. BMF-AI: BMF with user arguments and interaction cues. BMF-AIA: BMF with user arguments, interactions, and attributes; it corresponds to the full model presented earlier.

	MB	PMF	BMF	HFT	BMF-A	BMF-AI	BMF-AIA
Acc.	0.532	0.604 <sup>◊</sup>	0.607	0.642 <sup>◊</sup>	0.645	<b>0.703<sup>◊</sup></b>	<b>0.712<sup>◊</sup></b>
s.d.	0.015	0.012	0.012	0.011	0.012	0.008	<b>0.007</b>

Table 5: Micro-level stance prediction results, averaged across ten folds. <sup>◊</sup> The result is better than the method in the previous column at 5% significance level by McNemar’s test. s.d. refers to standard deviation.

We summarize the results in the following. (i.) MB performs poorly compared to other methods, since stances here are fairly balanced. (ii.) Both collaborative filtering approaches (PMF and BMF) significantly outperform the simple baseline. (iii.) BMF-A significantly outperforms BMF, at 5% significance level, meaning that text is helpful in modeling user stances; with user arguments, we are able to bring together issues that are similar, as evidenced by similar topic distributions. Meanwhile, BMF-A outperforms HFT by a very small margin; its chief advantage is its extensibility to multiple stances, which we have not tested. (iv.) Modeling the user interactions can further boost performance, as shown by BMF-AI outperforming BMF-A, at 5% significance level. (v.) By incorporating user attributes, the resulting model BMF-AIA achieves the best performance, with significance. This demonstrates the effectiveness of an integrated model of user arguments, interactions, and attributes.

**5.1.3 Cold-start Users.** For a cold-start user, although we don’t observe her past stances, our model can still profile her, using her attributes. Specifically, for a cold-start user  $u$ , we set factor deviations  $\delta_u = 0$ , and  $v_u = \frac{1}{Z_u} G^\top f_u$ . The user’s stance for an issue is:  $\arg \max_s \exp(v_u^\top v_{i,s})$ . Here  $G$  and  $v_{i,s}$  have been learnt from training data.

	MB	BMF-AI		
		+ Party	+ Religion	+ Both
Acc.	0.552	0.651	<b>0.662</b>	0.655
s.d.	0.037	0.028	0.029	<b>0.027</b>

Table 6: Micro-level stance prediction for cold-start users, averaged across ten folds. s.d. refers to standard deviation.

We compare our method with different types of attributes and a majority baseline. The results are presented in Table 6, which shows that our model

significantly outperforms the majority baseline, at 1% significance level (McNemar’s test). We also find that the *religion* attribute is a slightly (but not significantly) more important than *party* attribute in this task.

**5.2 Macro-Level Stance Prediction.** Recall that only 0.4% of the full set of user stances are observed in the data. We consider the task of predicting all the users’ stances on all issues; the aggregate of these gives a macro-level stance prediction. Using our model, we can predict any user’s stance on any issue in our data giving all the learnt variables in  $\rho$  according to equations 3.1 and 3.2. Specifically, we set  $\hat{r}_{u,i} = \arg \max_s \exp\{a_{u,i}^s\}$ , and the macro stance for an issue  $i$  is defined as:

$$(5.5) \quad \hat{n}_i^s = \sum_u \mathbf{I}(\hat{r}_{u,i} = s).$$

For a select number of issues with existing Gallup poll data,<sup>4</sup> we can evaluate the proportion of the total number of users holding each stance  $\frac{\hat{n}_i^s}{\sum_s \hat{n}_i^s}$  and compare them against the poll data.

The demographic of participants in CreateDebate are not representative of the larger population and hence we expect that these stance estimates do not match up with Gallup exactly.<sup>5</sup> Moreover, no one—neither us nor Gallup—can be certain that these estimates are accurate, due to difficulties in measuring public opinion such as sampling biases, truthfulness of responses, the way questions are framed, etc. (Of course, experts like those at Gallup have invested a great deal in techniques for overcoming those challenges.) Nevertheless, understanding where and how the results diverge may give us a sense of how the CreateDebate population is different from the population polled by Gallup.

We find there are multiple issues that are phrased differently but arguably mean the same thing, e.g., “Does God exist?” and “Is there a God?” We refer a group of such similar issues as *high-level issues*. We chose seven high-level issues with the most arguments that have corresponding Gallup polls. We select Gallup polls whose (*i.*) poll date is closest to the CreateDebate data collection date (April 2013) and (*ii.*) poll question is similar to all the issues in the high-level issue. For example, the high-level issue “gun control” contains three related issues: “Gun Control: Should we have it?”, “Should guns be banned?”, and “Should guns be

High-level issues	# issues	# users	Majority stance proportions
Believe God	3	274	54% Yes
Same sex marriage	3	91	71% Support
Abortion	4	86	58% Pro-life
2012 election	2	48	55% For Obama
Gun control	3	317	67% Against
Obamacare	2	13	62% Against
Death penalty	3	61	63% In favor

Table 7: Stance proportions of CreateDebate high-level issues used for macro-level stance predictions. The number of users and the majority stance in the table is aggregated across all similar issues from known user stances in the data.

banned in America?” and it corresponds to [37].<sup>6</sup>

For each issue  $i$  in CreateDebate, we know the stances of a small number of users and we can compute the proportions of users choosing the majority stance. We can also predict the proportions of the majority stance across the entire CreateDebate population with equation 5.5, and normalizing,  $\frac{\hat{n}_{i,s}}{\hat{n}_{i,s} + \hat{n}_{i,-s}}$ . Since we group similar issues together into high-level issues, stance proportions for similar issues are averaged to obtain stance proportions, which are presented in Table 7.

For these high-level issues, we identified Gallup poll results, from which we denote the ratios on both sides of the issue as  $c_{i,s}$  and  $c_{i,-s}$ . In Gallup polls, users are allowed to provide “no opinion” as an answer, meaning we have  $c_{i,s} + c_{i,-s} \neq 1$ . We ignore this small subset of polled users, and instead normalize the ratios to get for stance  $s$ :  $g_{i,s} = \frac{c_{i,s}}{c_{i,s} + c_{i,-s}}$ .

In Table 8, we present stance proportions predicted by Gallup polls, by our model, and by observed stances only. Treating Gallup polls results as “ground truth,” we define error of model-predicted stance proportions as  $\sum_{i,s} |g_{i,s} - \frac{\hat{n}_{i,s}}{\hat{n}_{i,s} + \hat{n}_{i,-s}}|$ , and error of observable stance proportions as  $\sum_{i,s} |g_{i,s} - \frac{n_{i,s}}{n_{i,s} + n_{i,-s}}|$ .

Overall, our results on macro-level stance prediction are encouraging. The mean absolute error of our predictions against Gallup is 0.07, much lower than 0.12 when using only observed user stances. By comparing with polls from Gallup, we find our methods tend to agree with or bring the stance prediction results closer to those online polls. For example, the Gallup poll results for “2012 election,” [39] has 49% of users vote for Obama. The CreateDebate corpus shows 55% of its users choosing against Obama. Our prediction model can share statistical strength, helping to estimate that

<sup>4</sup><http://www.gallup.com>

<sup>5</sup>CreateDebate reports that 89% of its users are between ages 20 and 40, 86.7% are from the United States, 85% are male, and 92% are single. These are based on self-reporting, but they give a strong sense that CreateDebate should not be taken as representative.

<sup>6</sup>The corresponding Gallup polls for these high-level issues are: believe God [38], abortion [39], 2012 elections [40], Obamacare [41], death penalty [42], gun control [37], and same sex marriage [43].

High-level issues	Gallup Pred. (Err.)		Maj. (Err.)
Believe God (Yes)	0.92	0.71 ( <b>0.21</b> )	0.54 (0.38)
Same-sex marriage (Support)	0.51	0.60 ( <b>0.09</b> )	0.71 (0.20)
2012 election (For Obama)	0.49	0.50 ( <b>0.01</b> )	0.55 (0.06)
Abortion (Pro-life)	0.52	0.54 ( <b>0.02</b> )	0.58 (0.06)
Gun control (Against)	0.75	0.69 ( <b>0.06</b> )	0.67 (0.08)
Death penalty (In Favor)	0.63	0.65 (0.02)	0.64 ( <b>0.01</b> )
Obamacare (Against)	0.53	0.64 (0.11)	0.61 ( <b>0.08</b> )

Table 8: Macro-level stance prediction results. The issues are sorted by difference in error between the two methods. in decreasing order. Prediction error refers to the difference between predicted proportions under our model and Gallup polls. Majority stances error refers to the difference between proportions of users with observed stances and Gallup polls.

around 50% of users support Obama, which is closer to the Gallup poll. This suggests that methods like ours might be used to complement traditional polling data collected via surveys. We find the high-level issue about belief in God has high prediction errors compared to others: Gallup [38] estimates 92% of users believe in God, while our model estimates 71%, with observable stances estimating only 54%. This shows that our model could make a prediction which is closer to the Gallup results, but still it cannot diverge the results too much from known stances. If our model is accurate, it suggests a dimension along which the population of CreateDebate users is quite different from Gallup’s sample.

## 6 Related Work

Existing approaches for stance prediction have focused on taking advantage of the availability of different sources of information.

Topic modeling approaches are useful for distilling texts into low dimensional topics, and using them for predicting user stances [44; 4; 1]. Somasundaran and Wiebe mined the web to augment existing data with learnt associations that are indicative of opinion stances in debates[2], and later focused on identifying stances in online debates by extracting useful linguistic features and making use of curated sentiment and argument lexicons [3]. Sentiment analysis has been used to infer interaction polarities and model the interplay between user interactions and stances [5; 6]. These methods require either rich textual contents that are present for every user, or use of an external, curated data source, which is not readily available in our problem setting. Data sparseness also makes standard supervised learning unsuitable for this task (see, e.g., [45]).

In this work, we consider collaborative filtering to alleviate the data sparseness problem. Collaborative filtering techniques can be readily applied to our user stance-issue matrix (see [46] for a survey). In our prob-

lem, we specifically consider PMF methods (proposed in [8]), because of its successful use in real-world problems [47; 48; 49; 50]. PMF has also been extended to incorporate social data. For instance, [49] extended it with social network information to perform social recommendation. Similarly, there are studies on recommendations that incorporate trust network information [51; 52; 53]. Our model bears similarity to these approaches in that we also consider social information extracted from user interactions. However, we incorporate user arguments to ground our model and learn intuitive explanations for the dimensions in our latent factors. We also incorporate user attributes into the model to help overcome cold-start, a notable problem with PMF.

We have compared our findings with publicly available poll data in macro-level stance prediction. Although inferring user opinions in the wider population is not our primary goal, we believe our method could play a role in using publicly available social media data to infer user opinions in the larger population [14; 15; 16; 54]. Our work takes this in a different direction, by studying online debates which have well structured user arguments, opinions, and interactions on diverse topics, to seek to predict user stances on a wider variety of social topics. To the best of our knowledge, this is the first study using debate forums for such analysis.

## 7 Conclusion

In this work, we studied the novel setting of stance prediction task in the online discussion forum CreateDebate, despite the low online participation rate of Internet users relating to any particular debate. We seek to predict user stances on a variety of topics; these methods might eventually complement traditional surveys. Our model brings together user arguments, interactions, and attributes into a collaborative filtering framework that exploits recently introduced fast inference methods. Experiments show promising results on both micro-level and macro-level stance prediction.

## Acknowledgments

This research is supported by the Singapore National Research Foundation under its International Research Centre @ Singapore Funding Initiative and administered by the IDM Programme Office, Media Development Authority (MDA), by an A\*STAR fellowship to Y. Sim, and by NSF CAREER IIS-IIS-1054319. We thank the reviewers for their valuable comments.

## References

- [1] W.-H. Lin, E. Xing, and A. Hauptmann, “A joint topic and perspective model for ideological discourse,” in *Machine Learning and Knowledge Discovery in Databases*. Springer, 2008, pp. 17–32.



- [2] S. Somasundaran and J. Wiebe, "Recognizing Stances in Online Debates," in *Proceedings of ACL-IJCNLP*, ser. ACL '09, 2009, pp. 226–234.
- [3] —, "Recognizing Stances in Ideological On-line Debates," in *NAACL Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, 2010, pp. 116–124.
- [4] M. Paul and R. Girju, "A Two-Dimensional Topic-Aspect Model for Discovering Multi-Faceted Topics," in *AAAI*, 2010.
- [5] M. Qiu and J. Jiang, "A Latent Variable Model for Viewpoint Discovery from Threaded Forum Posts," in *Proceedings of NAACL*, 2013, pp. 1031–1040.
- [6] M. Qiu, L. Yang, and J. Jiang, "Mining user relations from online discussions using sentiment analysis and probabilistic matrix factorization," in *Proceedings of the 2013 Conference of the NAACL*, Atlanta, Georgia, June 2013, pp. 401–410.
- [7] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry, "Using Collaborative Filtering to Weave an Information Tapestry," *Communications of the ACM – Special issue on Information Filtering*, vol. 35, no. 12, pp. 61–70, 1992.
- [8] A. Mnih and R. Salakhutdinov, "Probabilistic matrix factorization," in *NIPS*, 2007, pp. 1257–1264.
- [9] D. Agarwal and B.-C. Chen, "Regression-based latent factor models," in *Proceedings of KDD*, 2009, pp. 19–28.
- [10] J. McAuley and J. Leskovec, "Hidden Factors and Hidden Topics: Understanding Rating Dimensions with Review Text," in *Proceedings of RecSys*, 2013, pp. 165–172.
- [11] H. M. Wallach, "Topic Modeling: Beyond Bag-of-words," in *Proceedings of ICML '06*, 2006, pp. 977–984.
- [12] G. C. Wei and M. A. Tanner, "A monte carlo implementation of the em algorithm and the poor man's data augmentation algorithms," *Journal of the American Statistical Association*, vol. 85, no. 411, pp. 699–704, 1990.
- [13] L. Yao, D. M. Mimno, and A. McCallum, "Efficient methods for topic model inference on streaming document collections," in *KDD*, 2009, pp. 937–946.
- [14] B. O'Connor, R. Balasubramanian, B. R. Routledge, and N. A. Smith, "From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series," in *ICWSM*, 2010.
- [15] M. J. Paul and M. Dredze, "You Are What You Tweet: Analyzing Twitter for Public Health," in *ICWSM*, 2011.
- [16] A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Weppe, "Election Forecasts With Twitter: How 140 Characters Reflect the Political Landscape," *Social Science Computer Review*, 2010.
- [17] A. Abu-Jbara, P. Dasigi, M. Diab, and D. R. Radev, "Subgroup Detection in Ideological Discussions," in *Proceedings of ACL*, 2012, pp. 399–409.
- [18] A. Hassan, A. Abu-Jbara, and D. Radev, "Detecting Subgroups in Online Discussions by Modeling Positive and Negative Relations among Participants," in *EMNLP*, 2012, pp. 59–70.
- [19] P. Dasigi, W. Guo, and M. Diab, "Genre Independent Subgroup Detection in Online Discussion Threads: A Study of Implicit Attitude using Textual Latent Semantics," in *Proceedings of ACL (Volume 2: Short Papers)*, 2012, pp. 65–69.
- [20] Y. Lu, H. Wang, C. Zhai, and D. Roth, "Unsupervised Discovery of Opposing Opinion Networks from Forum Discussions," in *Proceedings of CIKM*, 2012, pp. 1642–1646.
- [21] A. Mukherjee and B. Liu, "Modeling Review Comments," in *Proceedings of ACL*, 2012, pp. 320–329.
- [22] —, "Mining Contentions from Discussions and Debates," in *Proceedings of KDD*, 2012, pp. 841–849.
- [23] M. Galley, K. McKeown, J. Hirschberg, and E. Shriberg, "Identifying Agreement and Disagreement in Conversational Speech: Use of Bayesian Networks to Model Pragmatic Dependencies," in *Proceedings of ACL*, 2004.
- [24] J. Andreas, S. Rosenthal, and K. McKeown, "Annotating Agreement and Disagreement in Threaded Discussion," in *Proceedings of LREC'12*, 2012.
- [25] M. Qiu, L. Yang, and J. Jiang, "Modeling interaction features for debate side clustering," in *22nd ACM CIKM'13*, 2013, pp. 873–878.
- [26] S. Gottipati, M. Qiu, Y. Sim, J. Jiang, and N. A. Smith, "Learning Topics and Positions from Debatepedia," in *Proceedings of EMNLP*, 2013, pp. 1858–1868.
- [27] O. Babko-Malaya, J. Pustejovsky, M. Verhagen, and A. Meyers, "Modeling debate within a scientific community," in *Social Intelligence and Technology (SOCIETY), 2013 International Conference on*, 2013, pp. 57–63.
- [28] Y. Sim, B. D. L. Acree, J. H. Gross, and N. A. Smith, "Measuring Ideological Proportions in Political Speeches," in *Proceedings of EMNLP*, October 2013, pp. 91–101.
- [29] L. Zhang, D. Agarwal, and B.-C. Chen, "Generalizing matrix factorization through flexible regression priors," in *Proceedings of RecSys '11*, 2011, pp. 13–20.
- [30] K. Chen, T. Chen, G. Zheng, O. Jin, E. Yao, and Y. Yu, "Collaborative personalized tweet recommendation," ser. SIGIR '12. ACM, 2012, pp. 661–670.
- [31] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.
- [32] A. Ahmed and E. P. Xing, "Staying Informed: Supervised and Semi-Supervised Multi-view Topical Analysis of Ideological Perspective," in *Proceedings of EMNLP*, 2010, pp. 1140–1150.
- [33] M. J. Paul, C. Zhai, and R. Girju, "Summarizing Contrastive Viewpoints in Opinionated Text," in *Proceedings of EMNLP*, 2010, pp. 66–76.
- [34] T. L. Griffiths and M. Steyvers, "Finding scientific topics," *Proceedings of the National Academy of Sciences*, vol. 101, pp. 5228–5235, 2004.
- [35] A. Ahmed, M. Aly, J. Gonzalez, S. Narayanamurthy, and A. J. Smola, "Scalable inference in latent variable models," in *WSDM '12*. ACM, 2012, pp. 123–132.
- [36] A. Q. Li, A. Ahmed, S. Ravi, and A. J. Smola, "Reducing the sampling complexity of topic models," in *Proceedings of the KDD '14*, 2014, pp. 891–900.
- [37] Gallup, "Remains divided passing stricter gun laws," November 2013, <http://www.gallup.com/poll/165563/remains-divided-passing-stricter-gun-laws.aspx>.
- [38] —, "Americans continue believe god," May 2011, <http://www.gallup.com/poll/147887/Americans-Continue-Believe-God.aspx>.
- [39] —, "Americans split along pro-choice pro-life lines," May 2011, <http://www.gallup.com/poll/147734/Americans-Split-Along-Pro-Choice-Pro-Life-Lines.aspx>.
- [40] —, "Romney obama gallup final election," November 2012, <http://www.gallup.com/poll/158519/romney-obama-gallup-final-election-survey.aspx>.
- [41] —, "Americans say health-law harmful helpful," May 2013, <http://www.gallup.com/poll/166793/americans-say-health-law-harmful-helpful.aspx>.
- [42] —, "Death penity support lowest years," October 2013, <http://www.gallup.com/poll/165626/death-penalty-support-lowest-years.aspx>.
- [43] —, "Sex marriage support solidifies above," May 2013, <http://www.gallup.com/poll/162398/sex-marriage-support-solidifies-above.aspx>.
- [44] Y. Fang, L. Si, N. Somasundaram, and Z. Yu, "Mining Contrastive Opinions on Political Texts Using Cross-perspective Topic Model," in *Proceedings of WSDM*, 2012, pp. 63–72.
- [45] W.-H. Lin, T. Wilson, J. Wiebe, and A. Hauptmann, "Which side are you on?: Identifying perspectives at the document and sentence levels," in *CoNLL-X '06*, 2006, pp. 109–116.
- [46] X. Su and T. M. Khoshgoftaar, "A survey of collaborative filtering techniques," *Adv. in Artif. Intell.*, vol. 2009, pp. 4:2–4:2, 2009.
- [47] Y. Koren, R. Bell, and C. Volinsky, "Matrix Factorization Techniques for Recommender Systems," *Computer*, vol. 42, no. 8, pp. 30–37, 2009.
- [48] R. M. Bell and Y. Koren, "Lessons from the Netflix Prize Challenge," *SIGKDD Explorations Newsletter*, vol. 9, no. 2, pp. 75–79, 2007.
- [49] H. Ma, H. Yang, M. R. Lyu, and I. King, "SoRec: Social Recommendation Using Probabilistic Matrix Factorization," in *Proceedings of CIKM '08*, 2008, pp. 931–940.
- [50] C. Wang and D. M. Blei, "Collaborative Topic Modeling for Recommending Scientific Articles," in *KDD*, 2011, pp. 448–456.
- [51] M. Jamali and M. Ester, "TrustWalker: A Random Walk Model for Combining Trust-based and Item-based Recommendation," in *Proceedings of KDD '09*, 2009, pp. 397–406.
- [52] —, "A matrix factorization technique with trust propagation for recommendation in social networks," in *Proceedings of RecSys*, 2010, pp. 135–142.
- [53] S.-H. Yang, A. J. Smola, B. Long, H. Zha, and Y. Chang, "Friend or Frenemy?: Predicting Signed Ties in Social Networks," in *Proceedings of SIGIR '12*, 2012, pp. 555–564.
- [54] M. Conover, J. Ratkiewicz, M. Francisco, B. Gonçalves, A. Flammini, and F. Menczer, "Political Polarization on Twitter," in *ICWSM*, 2011.